🏡 ImmoEliza — Team 3

> ### Get the Data
>
> - Scrape data from a property page → write to file
> - Scrape all property pages by calling `scrape_data()`
>   - ⚠️ This needs all the property links!

> ### Knowing What Data to Get
>
> - Get the property links from a search page → write to file
> - Get all the search pages

> ### Data Manipulation
>
> - Clean the data (remove duplicates)
> - Expect crashes → Write data to file

First Challenge

Getting past 50 pages of results

Build a new URL by changing the page number

First Idea: Filter by Province

> ❗ More than 1000 results for most provinces
>
> → Forward and reverse trick

- Still 4 provinces with more than 2000 results
  - Slice the data further? (subtypes, prices…)
    - → **Inelegant, messy, hard to automate…**

Better Idea: Filter by Zip Code

- Trick the site's API to give out all `{"zip_code": "city"}`
- Build a function to generate search page URLs
- Test it works
- → `All_Search_Pages.json`
- → `get_property_links_from_a_search_page()`
- → **ALL property links collected** ✅

Second Challenge: Cleaning Data

JSON with duplicates → New JSON without duplicates

Avoiding Crashes & Data Manipulation

- **Challenge:** Avoid data crashes
  - → Write each scraped line to a JSON
- **Manipulate data:**
  - → JSON to CSV converter

Ultimate Challenge

...solving the GitHub mess...

Solution

- Ask colleagues 🤝

Q&A

How do you organise your Repo?